

Transcrição

Neste passo, abordaremos a **variância**.

Como o desvio absoluto que vimos anteriormente, esta Medida de Dispersão também é uma média dos desvios, porém não é calculado somente com valores absolutos, e sim os **eleva ao quadrado** para não precisarem

Portanto, calcularemos a soma do quadrado dos desvios dividido pelo número total.

Teremos diferenças para calcularmos dados **populacionais** e **amostrais**. No caso da Estatística, geralmente trabalharemos com amostras.

Variância populacional

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

▼ Variância amostral

$$S^2 = \frac{1}{n - 1} \sum_{i=1}^n (X_i - \bar{X})^2$$

O primeiro cálculo é representado por "*sigma*" ao quadrado, sendo igual ao somatório dos desvios em relação à média populacional representada por "*mi*" e dividido por "*n*".

Já no segundo, dividiremos por "**n - 1**" conhecido como **Fator de Correção de Bessel**. Utilizando-o, teremos uma estimativa mais precisa do parâmetro populacional.

Usaremos o `df` feito anteriormente; pegaremos o `notas_fulano[]` com '`Desvio`', pois criaremos o desvio em relação à média elevado ao quadrado.

Colocaremos `Desvio` entre parênteses e aspas simples seguido de `^2`. Será igual `notas_fulano[]` com '`Desvio`' com aplicação do método Pandas `pow()` recebendo `2` para elevar ao quadrado.

```
notas_fulano['(Desvio)^2'] = notas_fulano['Desvio'].pow(2)
notas_fulano
```

Matérias	Fulano	Desvio	/Desvio/	(Desvio)^2
Matemática	8	0.285714	0.285714	0.81633
Português	10	2.285714	2.285714	5.224490
Inglês	4	-3.714286	3.714286	13.795918
Geografia	8	0.285714	0.285714	0.081633
História	6	-1.714286	1.714286	2.938776
Física	10	2.285714	2.285714	5.224490
Química	8	0.285714	0.285714	0.081633

Como resultado, veremos os valores das notas, o desvio, desvio absoluto e o desvio ao quadrado nas colunas.

Para calcularmos a variância amostral na célula seguinte, adicionaremos `notas_fulano[]` com o desvio ao quadrado. Se aplicarmos direto a média `mean()` como fizemos anteriormente, o sistema usará o "n" como d utilizar "n - 1".

Faremos o cálculo manualmente por meio da soma `.sum()` dividido por `len()` contendo `notas_fulano` entre parênteses para pegarmos o número de registros do conjunto e subtrairmos `1`.

```
notas_fulano['(Desvio)^2'].sum() / (len(notas_fulano) - 1)
```

O resultado será o valor `4.5714285714857` da variância das notas de Fulano.

Para calcularmos de maneira bem mais simples com Pandas, criaremos uma nova variável chamada `variancia` sendo igual a `notas_fulano[]` com as notas de `'Fulano'` seguido do método `var()`. Por fim, exibir

```
variancia = notas_fulano['Fulano'].var()
variancia
```

O valor do retorno será igual ao apresentado antes.

Como elevamos os desvios ao quadrado, também elevaremos as unidades da variância; suporemos que estamos calculando a `Altura` em metros de nosso dataset. Como resultado, obteremos em metros ao quadrado, o área, tornando mais difícil a interpretação dos dados.

Para resolver este problema, teremos o **desvio padrão** que abordaremos a seguir.